The Study on the Model of Effective Face Retrieval in Pedestrian Detection

Hua-Yang LI¹,a, Pan-Pan XIA²,b, Hui-Lian XU³,c

¹Jiangxi University of Finance and Economics, School of Software & Communication Engineering, China
²Jiangxi University of Finance and Economics, School of Software & Communication Engineering, China
³Jiangxi University of Finance and Economics, School of Foreign Languages, China

ª49834715@qq.com, b398368708@qq.com, c81036404@qq.com
²* Hua-Yang LI

Keywords: Effective Face Recognition, Symmetry Coefficient, Distance Coefficient.

Abstract. Pedestrian detection and retrieval plays an important role in intelligent video surveillance and automobile auxiliary driving system. Video surveillance creates mass monitored data among which a little effective information exists and could be hardly found. To settle the problem, based on pedestrian detection and face identification, the paper proposes a judgement and retrieval model of effective faces in pedestrian detection. The model helps to delete useless data and reserve the pedestrian data with effective faces, therefore, it provides data base for pedestrian image retrieval.

Introduction

It has achieved a lot in the researches of pedestrian detection with a great deal of manpower and funds due to the need for intelligent video surveillance. However, there exist difficulties in pedestrian detection and retrieval because features of pedestrians are changeable and complex in different situations. There is no universal algorithm with high performance for pedestrian detection and retrieval [1].

The histogram of oriented gradients (HOG) is a feature descriptor used in computer vision and image processing for the purpose of object detection. The feature is composed of statistical histograms of gradient information in the local region of the image [2,3,4]. To accelerate calculation speed, blocked histogram is calculated with integral images.

\[ h(x,y) = \sum_{s \leq x, t \leq y} h(x', y') \quad (1) \]

Eq. 1 defines integral image calculation. The integral image of \( M(x, y) \) is a gradient histogram of all the pixels in the upper left corner. \( h(x, y) \) refers to calculated integral image, \( h(x', y') \) is the gradient histogram of all the pixels in the upper left region. The concept of the integral histogram is shown in Fig. 1.

Fig. 1. A diagram of integral images

It is time consuming to view substantial data in surveillance videos. The purpose of the paper is to implement human retrieval in videos to construct database information for relevant departments to retrieve and query.
Our Work

It is essential to establish retrieval system to obtain relevant helpful images of pedestrians to greatly reduce the workload of some staff to settle problems. Images of people used to be retrieved should be the clearest facial images or shapes. Therefore, in the paper Effective Face refers to the clearest pictures of faces of people in a video. It is difficult to accurately judge the face posture with a single evaluation indicator[5,6]. To achieve accurate judgement of face images, the paper proposes two effective indicators, face area coefficient and face double symmetry coefficient. The effective coefficient of face symmetry is composed of distance symmetry and area symmetry of partial key points.

Moreover, the imaging area of an object is smaller in the distance than that nearby for a fixed surveillance device. Thus the overall area of a human face is an important parameter as well.

Distance symmetry of partial key points

The paper achieves face recognition for pedestrian images with Dlib toolkit which marks 68 key points. Suppose the symmetry coefficient of face key points is 1. If the face facing the lens squarely begins to move gradually towards the other side, its symmetry coefficient of face key points will decrease to 0 little by little. Not all parts of a face, such as the contour of the eye and mouth, own good distance symmetry features in the 68 key points marked in the face recognition. Therefore, a nose and the contour of the left and right side of a face are outlined as partial key points. The sequence number of the face contour key points is 1-17 and that of the nose is 28-36. The state of distance symmetry of selected key points is shown in Fig. 2.

Steps for calculating the distance symmetric coefficient of the partial key points is as follows.

Step 1 With the distance symmetric key point vector D, the contour feature vector D_2, and the nose feature vector D_2 are obtained. D_1, the feature vector of dimension 8, is obtained by recognizing the contour features D_1 according to symmetry.

\[
D_1 = \{P_1(x, y), P_2(x, y), P_3(x, y), \ldots, P_{17}(x, y)\} \tag{2}
\]

\[
D_2 = \{P_{28}(x, y), P_{29}(x, y), P_{30}(x, y), \ldots, P_{36}(x, y)\} \tag{3}
\]

\[
D_1' = \{(P_i(x, y), P_{17}(x, y)), \ldots, (P_i(x, y), P_{18-1}(x, y))\}, i = 1, 2, 3, \ldots, 8 \tag{4}
\]

Step 2 \(\bar{P}(x, y)\), the center of the nose, is calculated according to \(D_2\). \(\bar{P}(x, y)\) is regarded as a reference point when the contour symmetry is calculated.

\[
\bar{P}(x, y) = (\bar{x}, \bar{y}) = (\frac{\sum_{i=28}^{36} P_i^x}{9}, \frac{\sum_{i=28}^{36} P_i^y}{9}) \tag{5}
\]
Eq. 5 $\bar{x}$ is the abscissa of the center point, $\bar{y}$ is its ordinate. $P_i^x$ is the abscissa of the marked point $i$, $P_i^y$ is the ordinate of $i$.

Step 3 The corresponding distance symmetry coefficient $c_d^i$ is calculated for each reference point $\bar{P}(x, y)$ in $D_i$:

$$d_i = \sqrt{(x - P_i^x)^2 + (y - P_i^y)^2}$$
$$d_{18-i} = \sqrt{(x - P_{18-i}^x)^2 + (y - P_{18-i}^y)^2}$$
$$c_d^i = \frac{\min(d_i, d_{18-i})}{\max(d_i, d_{18-i})}$$

Eq. 6 $d_i$ is the distance from the key point $i$ to the center point $\bar{P}$. $d_{18-i}$ refers to the distance from the key point $18-i$ to the center point $\bar{P}$.

Step 4 Each key point is assigned with a weight and the distance symmetry coefficient $c_d$ is given after the result of step 3 is average-weighted

$$c_d = \frac{\sum_{i=1}^{N} c_d^i}{N} / 8$$

Known from the above steps, the closer the value of $c_d$ is to 1, the more symmetrical the face contour is to the nose. And the probability of having an effective face is greater. Due to a certain errors in marking key points of a face, the distance symmetry coefficient $c_d$ is not possibly to be equivalent to 1. Therefore $c_d$ is used to support an effective face in the process of ultimately determining which faces are effective or not.

**Area symmetry of partial key points**

The two eyes, highly symmetrical on a human face, are key reference factors in judging effective faces of pedestrians. The method of applying distance symmetry of key points to judge the features of a face is not suitable because key points in the eye region are highly concentrated without good distance symmetry. The paper puts forward the calculation method based on area symmetry of partial key points. First the left eye region and the right eye region are constructed with key points of the two eyes. Then the area of the two regions are calculated, with which their area ration is obtained to get area symmetric coefficient $c_s$. The closer $c_s$ is to 1, the more similar the area of the left and the right parts of eyes will be and the more squarely the face faces to the lens, from which judging effective face is more probable. The state of area symmetry of selected key points is shown in Fig. 3.

![Fig. 3. Selected Key Points in Area Symmetry](image)

(a) the key points of area are relatively symmetrical \hspace{1cm} (b) the key points of area are not symmetrical

Steps for calculating the area symmetric coefficient $c_s$ of the partial key points is as follows.

Step 1 Once $s$, a key point vector of the area symmetry, is decomposed, $s_1$ (a key point vector of the left eye) and $s_2$ (a key point vector of the right eye) are obtained. In calculation of the area of eye part,
two key points in the left region or the right region are included in the eye polygon, so the four key points are omitted when $S_1$ and $S_2$ are selected to calculate the area of eye region.

$$S_1 = P_i(x, y), i = 18,19,20,21,22,40,41,42,37$$

$$S_2 = P_i(x, y), i = 23,24,25,26,27,43,46,47,48$$

Step 2 $R_1$ and $R_2$ are respectively constructed. $R_1$ is the polygon region of the left eye, connected by points in $S_1$ and $R_2$ is the one of the right eye, connected by points in $S_2$.

Step 3 The area of $R_1$ and $R_2$ are respectively calculated. The area of $R_1$ is $A_1$, The area of $R_2$ is $A_2$.

Step 4 $c_s$, the area symmetry coefficient, is calculated.

$$c_s = \frac{\min(A_1, A_2)}{\max(A_1, A_2)}$$

A polygon is divided into several triangles whose area is calculated with the Helen formula. The sum of the area of the triangles is the polygon area.

**The coefficient of judging effective faces**

c, the coefficient of judging effective faces, is composed of the combination of distance symmetry coefficient and the area symmetry coefficient of partial key points.

$$c = f(c_d, c_s) = \frac{c_d + c_s}{2}$$

$$h(c, \theta) = c - \theta = \begin{cases} -1, & c - \theta \leq 0 \\ 1, & 0 \leq \theta \leq 1 \\ c - \theta > 0 \end{cases}$$

Eq. 12 $h(c, \theta)$ is a judging function and $\theta$ is a threshold parameter. Effective face and ineffective face are the result of judging faces in the paper. If $h(c, \theta) = -1$, the target image is judged to be ineffective face. If $h(c, \theta) = 1$, the target image is effective one.

**Effective face detection**

Based on effective face judgement, images are marked with facial geometric position features and image similarity is measured by the similar coefficient of double distance which is combined by weight Euclidean distance and average cosine distance. The steps of calculating facial geometric position features are as follows.

Step 1 Data of 68 key points are read and marked as $p$ which are divided to be $p_a$ and $p_b$. $p_a$ is the feature points of the eye, contour and the mouth. $p_b$ is feature points of a nose.

$$P = (P_1(x, y), P_2(x, y), P_3(x, y), \ldots, P_{68}(x, y))$$

$$p_a = P - p_b$$

$$p_b = (P_{28}(x, y), P_{29}(x, y), P_{30}(x, y), \ldots, P_{36}(x, y))$$

Step 2 $\bar{P}(\bar{x}, \bar{y})$, the facial center, is calculated by making reference to all the key points in $p_b$.

$$\bar{P}(\bar{x}, \bar{y}) = (\frac{\sum_{i=28}^{36} P_{i}'x}{9}, \frac{\sum_{i=28}^{36} P_{i}'y}{9})$$

Eq. 13 $\bar{x}$ is the abscissa of the center point, and $\bar{y}$ is its ordinate. $P_{i}'x$ is the abscissa of the marked point $i$ and $P_{i}'y$ is $i$’s ordinate.

Step 3 The maximum distance $d$ is obtained by calculating the distance $d_i$ from the feature point to the center point.
Eq. 14 $\bar{x}$ is the abscissa of the center point, and $\bar{y}$ is its ordinate. $P_i^x$ is the abscissa of the key point $i$ and $P_i^y$ is its ordinate, $d_i$ presents the distance from the key point $i$ to the center point, $d$ is the maximum value of distance.

Step 4 $d$ is the denominator of distance normalization. The distance of all feature points is normalized to the same unit length, and the 59 normalized features are constructed as a facial geometric feature vector $F$.

$$
F = \left(\frac{d_1}{d}, \frac{d_2}{d}, \frac{d_3}{d}, \ldots, \frac{d_{59}}{d}\right)
$$

(19)

### Similarity calculation Weighted Euclidean distance

First, the similarity coefficient of two images is defined to be $D$. Then $WD_e$ and $MD_e$ are obtained by calculating similar distances of facial geometric position features with the improved method. Where, the weighted Euclidean distance with the value range $[0, +\infty)$, the smaller the value is, the more similar the two images are. The mean cosine distance with the value range $[-1, 1]$, the bigger the value is, the more similar the two images are. To combine the two coefficients, the similarity coefficients are defined as follows.

$$
D = \frac{1}{1 + WD_e} + MD_e
$$

(20)

Eq. 20 the value range of $\frac{1}{1 + WD_e}$ is $[0, 1]$ and the value range of $MD_e$ is $[-1, +1]$. Thus the value range of the similarity coefficient $D$ is $[-1, 2]$, which shows the bigger the value is, the more similar the two images will be.

### Analysis of the experiment

The hardware applied in the experiment is as follows.

<table>
<thead>
<tr>
<th>Table 1 Experimental Hardware Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Operation System</strong></td>
</tr>
<tr>
<td>Memory(RAM)</td>
</tr>
<tr>
<td>Processor</td>
</tr>
<tr>
<td>Hard Disk</td>
</tr>
<tr>
<td>Toolkit</td>
</tr>
<tr>
<td>Development Language</td>
</tr>
</tbody>
</table>

### Analysis of pedestrian detection

INRIA Database of pedestrian detection and recorded video data are applied in the experiment. INRIA Database is used to train pedestrian detection classifier. With the trained classifier pedestrian recorded in the experimental data are detected. INRIA Database contains test and training data sets. In test data, 288 samples are positive and 453 samples negative. In testing data, 613 samples are positive and 1218 samples negative. The experimental data are practical in actual application situations because the data acquisition is closely influenced by the complex background, light conditions and human gestures.
Table 2 Comparison of detection performance with different HOG feature methods

<table>
<thead>
<tr>
<th>Feature method</th>
<th>Recognition rate</th>
<th>Error detection rate</th>
<th>Average detection time</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOG feature</td>
<td>72.1%</td>
<td>40%</td>
<td>2.3s</td>
</tr>
<tr>
<td>HOG features of 3D linear interpolation</td>
<td>80.2%</td>
<td>27%</td>
<td>7s</td>
</tr>
<tr>
<td>Integral histogram accelerated HOG feature</td>
<td>70.6%</td>
<td>43%</td>
<td>0.38s</td>
</tr>
</tbody>
</table>

In the analysis of the detection result, higher recognition rate exists in the detection with 3D linear interpolation and faster detection speed appears with integral images. A pedestrian target will appear in several continuous frames in a video file and in the stage of judging effective faces some images of pedestrians will be filtered, so pedestrian can be well detected in a video with Integral Histogram as shown in Picture 1.

**Analysis of effective face judgement**

Based on pedestrian detection, effective faces are judged in the pedestrian data. In the process face key points mark and

![Fig. 4. Effect detected picture with Integral histogram accelerated HOG feature](image)

![Fig. 5. Pedestrian images](image)

![Fig. 6. Corresponding images of face key points](image)
Effective coefficients of images are calculated according to marked result. Judgement coefficients of 6 images in Fig. 6 are shown in Table 3.

<table>
<thead>
<tr>
<th>No.</th>
<th>Distance symmetry coefficient of partial key points</th>
<th>Area symmetry coefficient of partial key points</th>
<th>Double symmetry coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>A2</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>A3</td>
<td>0.598937384795</td>
<td>0.599156118143</td>
<td>0.599046751469</td>
</tr>
<tr>
<td>A4</td>
<td>0.742306122578</td>
<td>0.754166666667</td>
<td>0.7482363946225</td>
</tr>
<tr>
<td>A5</td>
<td>0.896432752106</td>
<td>0.884955752212</td>
<td>0.890694252159</td>
</tr>
<tr>
<td>A6</td>
<td>0.938131843406</td>
<td>0.911016949153</td>
<td>0.9245743962795</td>
</tr>
</tbody>
</table>

Known from the experimental result, distance symmetry coefficient and area symmetry coefficient of partial key points are much closer to 1 when pedestrians face the lens more squarely.

Summary

The paper proposes a model of judging and retrieving effective faces based on the technology of pedestrian detection and face recognition. The model optimizes the feature extraction method with integral images by detecting pedestrians in videos and retrieve effective faces with double symmetry coefficient to obtain more effective pedestrian data.

Acknowledgement

The paper is sponsored by visiting Scholar Project of Faculty Development of Jiangxi Province.

References


